# Homework 4 - Markov chains

**1) Transforming a random process into a Markov chain.** Consider the random process $X_{\mathbb{N}} = X_0, X_1, \ldots, X_n, \ldots$ with state space $S = \{0, 1, 2\}$. Suppose

$$P\left[X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\right] = \left\{ \begin{array}{ll} P_{ij}^e, & n \text{ even} \\ P_{ij}^o, & n \text{ odd} \end{array} \right.$$

where $\sum_{j=0}^2 P_{ij}^e = \sum_{j=0}^2 P_{ij}^o = 1$, $i \in S$. Is $X_{\mathbb{N}}$ a Markov chain (MC)? Explain. If not, then show how, by augmenting the state space, we may transform it in a MC. Give the resulting transition probability matrix.

**2) Coin flips.** Suppose that coin 1 has probability 0.7 of coming up heads, and coin 2 has probability 0.6 of coming up heads. If the coin flipped today comes up heads, the we select coin 1 to flip tomorrow, and if it comes up tails, then we select coin 2 to flip tomorrow. If the coin initially flipped is equally likely to be coin 1 or coin 2, then what is the probability that the coin flipped on the third day after the initial flip is coin 1? Suppose that coin flipped on Monday comes up heads. What is the probability that the coin flipped on Friday of the same week also comes up heads?

**3) Positive transition matrices remain positive.** Let $\mathbf{P}$ be the transition probability matrix of a MC. Show that if $\mathbf{P}^r$ has all positive entries for some positive integer $r$, then so does $\mathbf{P}^n$ for all $n \geq r$.

**4) Communication classes.** Specify the communication classes of the following MCs, and determine whether they are transient or recurrent. Are the MCs irreducible?

$$\mathbf{P}_1 = \left( \begin{array}{ccc} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{array} \right), \quad \mathbf{P}_2 = \left( \begin{array}{cccc} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array} \right),$$

$$\mathbf{P}_3 = \left( \begin{array}{ccccc} 1/2 & 0 & 1/2 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{array} \right), \quad \mathbf{P}_4 = \left( \begin{array}{ccccc} 1/4 & 3/4 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{array} \right).$$

**5) Multi-step, unconditional, and joint MC probabilities.** Suppose that $X_{\mathbb{N}} = X_0, X_1, \ldots, X_n, \ldots$ is a MC with state space $S = \{1, 2\}$, transition probability matrix

$$\mathbf{P} = \left( \begin{array}{cc} 1/5 & 4/5 \\ 2/5 & 3/5 \end{array} \right)$$

and initial distribution $P[X_0 = 1] = 3/4$ and $P[X_0 = 2] = 1/4$. You are asked to calculate $P\left[X_3 = 1 \mid X_1 = 2\right]$, $P\left[X_3 = 1 \mid X_2 = 1, X_1 = 1, X_0 = 2\right]$, $P[X_2 = 2]$, $P[X_0 = 1, X_2 = 1]$ and $\mathbb{E}[X_2]$.

**6) Something the Markov property does not imply.** Suppose that $X_{\mathbb{N}} = X_0, X_1, \ldots, X_n, \ldots$ is a MC, $j, i_{n-1}, \ldots, i_1, i_0$ are elements of the state space $S$, and $A$ is a subset of the state space. Give an example for which

$$P\left[X_{n+1} = j \mid X_n \in A, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0\right] \neq P\left[X_{n+1} = j \mid X_n \in A\right].$$

(Hint: construct a MC where state $i_{n-1}$ communicates with a single state $i_n \in A$, and let $A = S$.)

**7) A branching process.** Mitochondrial DNA is passed from mother to children without genetic contribution from the father. All the variability in mitochondrial DNA is due to random mutations accumulated over time. Using estimates of the mutation rate and differences on mitochondrial DNA between humans it becomes possible to estimate the time at which groups became distinct populations. We are not computing these times here but a few facts about a MC that models the propagation of mitochondrial DNA. Since male mitochondrial DNA is not passed to children, it suffices to focus on female descendants of females, i.e., from mothers to daughters to daughters of daughters and so on.

Let us start denoting as $X_n$ the total number of women in the $n$-th generation and as $X_{rn}$ the total number whose mitochondrial DNA is of type $r$. Assign indexes $i = 1, 2, \ldots, X_n$ to the $n$-th generation individual. Independently of time and her type $r$ woman $i$ has $D_i$ daughters with probability distribution

$$P[D_i = j] = p_j, \quad \text{for all } i. \tag{1}$$

In most cases daughters' types coincide with their mothers' types. Once in a while, though, a mother of type $r$ bears a daughter of a different type $s$. When this happens, we assume that type $s$ is novel, i.e., different from other types present in the population. This is a reasonable assumption because mutations are rare and can happen in a large number of genes. The probability of the same mutation appearing twice can thus be discounted. The appearance of a new group happens with probability $q$. To simplify calculations we still count the mother as being of type $r$, and start counting type $s$ on the daughters' generation. We can then separate the probability of bearing $D_{ir}$ daughters of type $r$ or $D_{is}$ daughters of type $s$ as

$$P[D_{ir} = j] = (1 - q)p_j, \qquad P[D_{is} = j] = qp_j \tag{2}$$

where in the above equation the woman $i$ is assumed to be of type $r$. The first probability accounts for the case when no mutations occur. The second one accounts for the daughters of a woman in which the mutation first arises. If at generation $n$ none of the $X_{rn}$ type $r$ women bears any daughter of type $r$, we say type $r$ goes extinct and accordingly $X_{rk} = 0$, for all $k \geq n + 1$.

For future use define $\nu$ as the expected value of the number of daughters, and $\nu_r$ as the expected number of daughters that share their mother's type, i.e.,

$$\nu := \mathbb{E}[D_i] = \sum_{j=1}^{\infty} jp_j, \qquad \nu_r := \mathbb{E}[D_{ir}] = (1 - q)\sum_{j=1}^{\infty} jp_j. \tag{3}$$

*A) Is the count of women a Markov chain?* Consider first the total number of women $X_n$. Is the process $X_{\mathbb{N}}$ a MC? If the answer to this question is positive, what are the transition probabilities $P_{ij} := P[X_{n+1} = j \mid X_n = i]$ from states $X_n = 0$ and $X_n = 1$, i.e, for $i = 0$, $i = 1$ and any $j$? What are the transition probabilities into state $X_n = 0$, i.e., for $j = 0$ and any $i$? Is the probability $P_{ii} := P[X_{n+1} = i \mid X_n = i]$ of a state transitioning into itself strictly positive? Is this MC recurrent?

*B) Is the count of women classified by mitochondrial DNA type a Markov chain?* Consider now the count of women $X_{rn}$ with mitochondrial DNA of type $r$. As defined the process $X_{r\mathbb{N}}$ is *not* a MC. Why? (Hint: state $X_{rn} = 0$ can either represent that a mutation onto type $r$ has yet to occur, or, that the type $r$ population is extinct. Hence, do transition probabilities $P[X_{r,n+1} = j \mid X_{rn} = i]$ depend only on $i = 0$ and $j$?) Suppose we are given the information that at some time $n$, $X_{rn} > 0$. Define the random process $X_{r,n:\infty} := X_{rn}, X_{r,n+1}, \ldots$ that starts at time $n$ with the information that $X_{rn} > 0$. This process is a MC. Why? What are the transition probabilities $P_{ij} := P[X_{r,n+1} = j \mid X_{rn} = i]$ from states $X_{rn} = 0$ and $X_{rn} = 1$, i.e, for $i = 0$, $i = 1$ and any $j$? What are the transition probabilities into state $X_{rn} = 0$, i.e., for $j = 0$ and any $i$? Is the probability $P_{ii} := P[X_{r,n+1} = i \mid X_{rn} = i]$ of a state transitioning into itself strictly positive? Is this MC recurrent?

*C) System simulation.* Write a simulation of this stochastic system. You can model the number of children as Poisson with mean $\lambda = 1.05$ (i.e., $p_j = e^{-1.05}(1.05)^j/j!$), which is half the fertility rate in the United States. Your function should take as inputs $\lambda$, the rate of mutation $q$, the duration of the experiment $n_{\max}$ (in number of generations), and the initial number of individuals $X_0$ all having different mitochondrial DNA types. The outputs should be a matrix with entries $X_{rn}$, and two vectors of length $n_{\max}$ indicating: i) the total number of mitochondrial DNA types by generation $n$ (including those that went extinct); and ii) the total number of extinct types by generation $n$. Hand in your code.
**Optional:** If you are up for a challenge, you can approximate probabilities from the following distribution of the number of children ever born to women in the age group 40-44[1]:

| Number of children | Percentage | Number of children | Percentage |
|---|---|---|---|
| 0 | 0.179 | 1 | 0.174 |
| 2 | 0.354 | 3 | 0.189 |
| 4 | 0.068 | 5,6 | 0.028 |
| > 7 | 0.008 | | |

If you decide to use the data in this table, notice that the above distribution is for *all* children, male and female, and that you are interested in girls only.

*D) Simulation tests one.* Run a simulation with $q = 10^{-2}$ as the rate of mutation, initialize generation 0 with $X_0 = 100$ individuals all having different mitochondrial DNA types. Run for $n_{\max} = 50$ generations – approximately 1,000 years at 20 years per generation. Show a plot for the number of women in each type as a function of generation index. Plot

---

[1]US Census Bureau, "Distribution of Women by Average Number of Children Ever Born, by Race, Marital Status, and Age," June 2002

your two output vectors to show the number of types present and accumulated number of extinct types as as a function of generation index. For $n = 50$, show a histogram of the number of individuals per type (including those extinct types).

*E) Simulation tests two.* Repeat Part *D* with rate of mutation $q = 0$ and $X_0 = 400$ individuals of different types, so here the number of types present is constant and equal to $400$ for all $n \geq 0$. This tells you about the chances of your direct female line surviving into the next 10 centuries. Notice that most of the types go extinct, a few have a moderate number of individuals and 1 or 2 have a large number of individuals. This means that far into the future, most of your direct female lines will be extinct, but one of you will have a very large number of survivors. But who among you is the one surviving into the future is determined by chance.

*F) Expected value of the number of direct line female descendants.* The number of individuals in the $(n+1)$-st generation can be written in terms of the corresponding number $X_n$ for the $n$-th generation and the numbers of daughters $D_i$ of each individual. The expression is

$$X_{n+1} = \sum_{i=1}^{X_n} D_i. \tag{4}$$

Use (4) to prove that if the number of women in the first generation ($n = 0$) is $X_0$ then the expected number of female individuals in the $n$-th generation is

$$\mu_n = \mathbb{E}[X_n] = X_0 \nu^n. \tag{5}$$

Compare this expected value with the number of individuals as a function of time that you obtained in Parts *D* and *E*. If they're similar explain the similarity. If they're not similar explain the differences. Likewise, prove that the expected number of descendants of a generation 0 individual sharing her mitochondrial type is

$$\mu_{rn} = \mathbb{E}[X_{rn}] = \nu_r^n = (1 - q)^n \nu^n. \tag{6}$$

*G) Extinction in probability.* Show that if $\nu_r < 1$ then type $r$ goes extinct in probability independently of the number of individuals in the original generation, i.e,

$$\lim_{n \to \infty} \mathbf{P}[X_{rn} = 0] = 1. \tag{7}$$

*H) Probability of eventual extinction ($q = 0$).* Fix $q = 0$. Denote as $P_e(j)$ the probability of eventual extinction of mitochondrial type $r$, when the number of generation 0 type $r$ individuals is $X_{r0} = j$, i.e.,

$$P_e(j) := \lim_{n \to \infty} \mathbf{P}\left[X_{rn} = 0 \,\middle|\, X_{r0} = j\right]. \tag{8}$$

To explain this definition consider a single individual at time 0, $P_e(1)$ is the probability that her direct female line eventually dies out. We already showed that for $\nu < 1$, then $P_e(j) = 1$ independent of $j$. When $\nu > 1$ is turns out that $P_e(j) < 1$, and the equation determining $P_e(1)$ can be obtained by conditioning on the number of daughters in the first generation, and using the law of total probability. Show that $P_e(1)$ is a solution of the following equation

$$P_e(1) = \sum_{j=1}^{\infty} p_j \, P_e^j(1). \tag{9}$$

Argue that in general $P_e(j) = \left[P_e(1)\right]^j$.